

Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from *Clostridium thermocellum*

Edward J. TAYLOR*, Nicola L. SMITH†, Johan P. TURKENBURG*, Simone D'SOUZA*, Harry J. GILBERT† and Gideon J. DAVIES*¹

*York Structural Biology Laboratory, Department of Chemistry, University of York, York YO10 5YW, U.K., and †Institute for Cell and Molecular Biosciences, University of Newcastle upon Tyne, Newcastle upon Tyne NE2 4HH, U.K.

The digestion of the plant cell wall requires the concerted action of a diverse repertoire of enzyme activities. An important component of these hydrolase consortia are arabinofuranosidases, which release L-arabinofuranose moieties from a range of plant structural polysaccharides. The anaerobic bacterium *Clostridium thermocellum*, a highly efficient plant cell wall degrader, possesses a single α -L-arabinofuranosidase (EC 3.2.1.55), CtAraf51A, located in GH51 (glycoside hydrolase family 51). The crystal structure of the enzyme has been solved in native form and in 'Michaelis' complexes with both α -1,5-linked arabinotriose and α -1,3 arabinoxylobiose, both forming a hexamer in the asymmetric unit. Kinetic studies reveal that CtAraf51A, in contrast with well-characterized GH51 enzymes including the *Cellvibrio japonicus* enzyme

[Beylot, McKie, Voragen, Doeswijk-Voragen and Gilbert (2001) Biochem. J. 358, 607–614], catalyses the hydrolysis of α -1,5-linked arabino-oligosaccharides and the α -1,3 arabinosyl side chain decorations of xylan with equal efficiency. The paucity of direct hydrogen bonds with the aglycone moiety and the flexible conformation adopted by Trp¹⁷⁸, which stacks against the sugar at the +1 subsite, provide a structural explanation for the plasticity in substrate specificity displayed by the clostridial arabinofuranosidase.

Key words: arabinan, arabinofuranosidase, *Clostridium thermocellum*, glycosidase, ligand specificity, xylan.

INTRODUCTION

The plant cell wall is a complex composite of structural polysaccharides that represents the most abundant source of organic molecules in the biosphere. The annual recycling of 10¹¹ tons of plant structural polysaccharides is an important biological process that is integral to the carbon cycle. Indeed, enzymes, such as xylanases, cellulases, pectinases and mannanases, which play a key role in plant cell wall recycling, are already used in biotechnological processes in the paper/pulp, fruit juice, detergent and textile sectors (some examples include [1,2]). The exploitation of the plant cell wall as an environmentally renewable energy source, estimated to equate to 600 billion barrels of oil, represents the major industrial application of this complex macromolecular structure. In this process, the plant cell wall is hydrolysed by microbial enzymes to its composite sugars, which are then used as cheap substrates in industrial fermentations to generate 'biofuel' ethanol [3]. The complete saccharification of the plant cell wall requires an extensive repertoire of hydrolytic enzymes that include both the endo-acting enzymes described above and an extensive range of exo-acting glycosidases.

Pectins and hemicelluloses, which comprise the major matrix polysaccharides in the plant cell wall, are extensively decorated with arabinofuranose moieties. Exo-acting α -L-arabinofuranosidases hydrolyse terminal and side chain arabinofuranosides in arabinan (α -1,5-linked arabinofuranosides), and the arabinosyl decorations within arabinogalactan and arabinoxylan (decorated at the C-2 and C-3 position) [4] (Figure 1). Arabinofuranosidases are located in the sequence-based GHs (glycoside hydrolase families; defined in [5,6] and recently reviewed in [7]) GH3, GH43, GH51, GH54 and GH62. The different families display widely varying specificity, although the paucity of characterized

members makes generalization difficult. For example, GH51 and GH54 arabinofuranosidases have been shown to remove both α -1,2 and α -1,3 arabinofuranosyl moieties from arabinan and xylans [8], the GH43 enzymes hydrolyse α -1,5-linked arabinofuranoligosaccharides, while GH62 glycoside hydrolases display absolute specificity for arabinoxylans [9–13]. At the structural level, there is a single α -L-arabinofuranosidase structure, the *Geobacillus stearothermophilus* enzyme from family GH51 [14]. This structure revealed a functional hexamer with each constituent monomer displaying a 'clan GH-A' [15] catalytic (β/α)₈ domain linked to a C-terminal 12-stranded β -barrel.

Clostridium thermocellum is an anaerobic thermophilic soil bacterium that displays a spectrum of plant cell wall-degrading enzymes. The vast majority of the cell wall-degrading glycosidases from this organism are arranged on to a mega-Dalton supramolecular assembly termed the cellulosome (recently reviewed in [16]). The bacterium, however, also synthesizes intracellular glycoside hydrolases whose function is to hydrolyse oligosaccharides released from the plant cell wall by the cellulosome. One such enzyme is the GH51 arabinofuranosidase, hereafter termed CtAraf51A. Here, we report the characterization of CtAraf51A, cloned from genomic DNA, and its three-dimensional structures in native form and in 'Michaelis' complexes, of an inactive mutant form of CtAraf51A, with A3 (α -1,5-linked arabinotriose) and AX2 (α -1,3-linked arabinoside of xylobiose; derived from xylanase breakdown of arabinoxylan). It is shown that CtAraf51A is equally efficient at removing the side chain decoration of AX2 (xylobiose in which the non-reducing xylose is decorated with arabinofuranose at O-3) as it is in hydrolysing the backbone of A3, an unusual specificity which is discussed in light of the three-dimensional complexes of the enzyme.

Abbreviations used: A3, α -1,5-linked arabinotriose; AX2, α -1,3-linked arabinoside of xylobiose; arabinan, α -1,5-linked arabinofuranosides; GH51, glycoside hydrolase family 51; LB broth, Luria–Bertani broth; PNP-Araf, 4-nitrophenyl- α -L-arabinofuranoside.

¹ To whom correspondence should be addressed (email davies@ysbl.york.ac.uk).

The co-ordinates of CtAraf5 have been submitted to the RSCB PDB under PDB identifiers 2C8N and 2C7F.

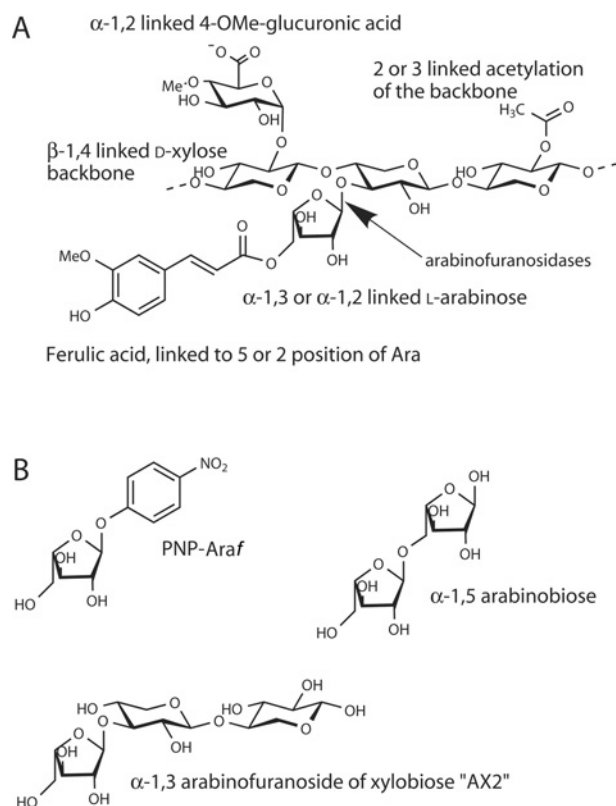


Figure 1 The enzymatic breakdown of xylan (A) with diverse substrates used in the characterization of *CtAraf51* (B)

EXPERIMENTAL

Expression and purification of *CtAraf51* and mutant derivatives

Escherichia coli strains NovoBlue and BL21 DE3 (Novagen) were cultured at 37°C in LB (Luria–Bertani) broth. Media were supplemented with 50 mg/l kanamycin.

The open reading frame encoding *CtAraf51A* was amplified from a *Cl. thermocellum* ATCC 27405 genomic DNA template by PCR using the forward and reverse primers: CACCACCACCA-CATGAAAAAGCCAGAATGACCGTTGACAAAGATTATA-AAATTGCC and GAGGAGAAGGCGCGTTATTATTACCTA-TCCGAATTACATTCCAAGAGGCTCTGCGAAG respectively. These were designed to incorporate LIC (ligation-independent cloning) and compatible sequence and the resultant 1542 bp product annealed into vector ysbLIC pET28a to produce the expression construct pGH51. This construct sequence was confirmed by DNA sequencing (MWG Biotech) and found to be identical with the known genomic sequence for the *Cl. thermocellum araf51a* gene (> gil29341675). The pGH51 expression construct provided a template for site-directed mutagenesis to produce the acid/base D175A mutant. Specific primers: forward: GGTGTCTTGGCA-ATGAAATGGACGGTCCG reverse: CGGACCGTCCATTTC-ATTGCCAAGACACC were used in conjunction with the 'Quick Change' mutagenesis kit (Stratagene). Conformation of correct mutagenesis was obtained by DNA sequencing.

CtAraf51A protein was produced in *E. coli* BL21 DE3 cells harbouring the appropriate recombinant plasmid, cultured in LB broth containing kanamycin at 37°C. Cells were grown to mid-exponential phase [A_{350} (absorbance) of 0.7], at which point isopropyl β -D-thiogalactopyranoside was added to a final concentration of 1 mM and the cultures were incubated for a further 14 h

at 16°C. The cells were harvested by centrifugation and His₆-tagged recombinant proteins were purified from cell-free extracts by IMAC (immobilized metal ion affinity chromatography) following a 15 min heat denaturation step at 70°C. For crystallization, *CtAraf51A* was further purified by gel filtration.

Enzyme assays

Assays were carried out using either PNP-Araf (4-nitrophenyl- α -L-arabinofuranoside; Sigma), 3'-O-arabinofuranosyl-xylobiose, 3''-O-arabinofuranosyl-xylotriose or 3'-O-arabinofuranosyl-xylotriose and the Megazyme International products arabinobiose, wheat arabinoxylan, sugar beet arabinan and linear arabinan. The arabinoxyl-oligosaccharides were prepared by treating arabinoxylan with a GH10 xylanase either partially, to generate 3''-O-arabinofuranosyl-xylotriose, or to completion to produce 3'-O-arabinofuranosyl-xylobiose and 3'-O-arabinofuranosyl-xylotriose. The oligosaccharides were purified by two rounds of gel filtration using P2 columns (Bio-Rad) as described by Proctor et al. [17]. The structure of the molecules were verified by high-performance anion-exchange chromatography analysis of the products generated by digestion with the GH51 arabinofuranosidase, a GH39 β -xylosidase, which removes undecorated xylose residues from the non-reducing ends of xylo-oligosaccharides, or acid hydrolysis. To assay for PNP-Arafase activity, the standard reaction conditions comprised 50 mM sodium dihydrogen phosphate brought to pH 7.0 with NaOH, containing 1 mg/ml BSA and substrate concentrations that ranged from 0.5 to 2 times the K_m value. The reaction was incubated with the enzyme (ranging from 4 nM to 8 μ M) at 37°C and the increase in A_{400} was measured against time. The amount of product generated was calculated using a molar absorption coefficient (ϵ), at 400 nm wavelength for 4-nitrophenol at pH 7.0, of 10 500 M⁻¹ · cm⁻¹. To measure the activity of the arabinofuranosidase against polysaccharides and oligosaccharides, the reaction was supplemented with 1 mM NAD⁺ and 50 μ M galactose dehydrogenase (Megazyme International) and the release of arabinose was monitored at A_{340} to detect NADH production. To determine the temperature optimum of the arabinofuranosidase, the enzyme was assayed at various temperatures using 1 mM PNP-Araf as the substrate. Thermostability was assessed by incubating the enzyme at various temperatures for up to 60 min. Aliquots were removed at regular time intervals and assayed for PNP-Arafase activity at 37°C using substrate at 1 mM.

Crystallization and data statistics

Pure protein as judged by SDS/PAGE was concentrated to 28 mg/ml and buffer exchanged into 20 mM Hepes using a Vivaspin 10 kDa cut-off concentrator. *CtAraf51A* crystallization conditions were screened using the hanging drop vapour diffusion method together with the Hampton Crystal screen, Crystal screen 2 and Hampton PEG/Ion screen (Hampton Research, Aliso Viejo, CA, U.S.A.). Drops containing 1 μ l of protein were mixed together with 1 or 2 μ l of the mother liquor. Initial crystals were found to grow in Hampton screen 1, condition number 7. This condition was optimized further to improve crystal quality, resulting in optimized conditions of 5 M sodium acetate, 0.1 M sodium cacodylate (pH 6.5) and 5% (v/v) dioxane. A cryoprotectant solution was produced through the addition of 20% (v/v) ethylene glycol to the composition of the mother liquor. The crystals were harvested in rayon fibre loops, bathed in cryoprotectant solution prior to flash freezing in liquid nitrogen. Ligand complexes were obtained by soaking wild-type or mutant crystals in 1 μ l drops of mother liquor, and grains of ligand

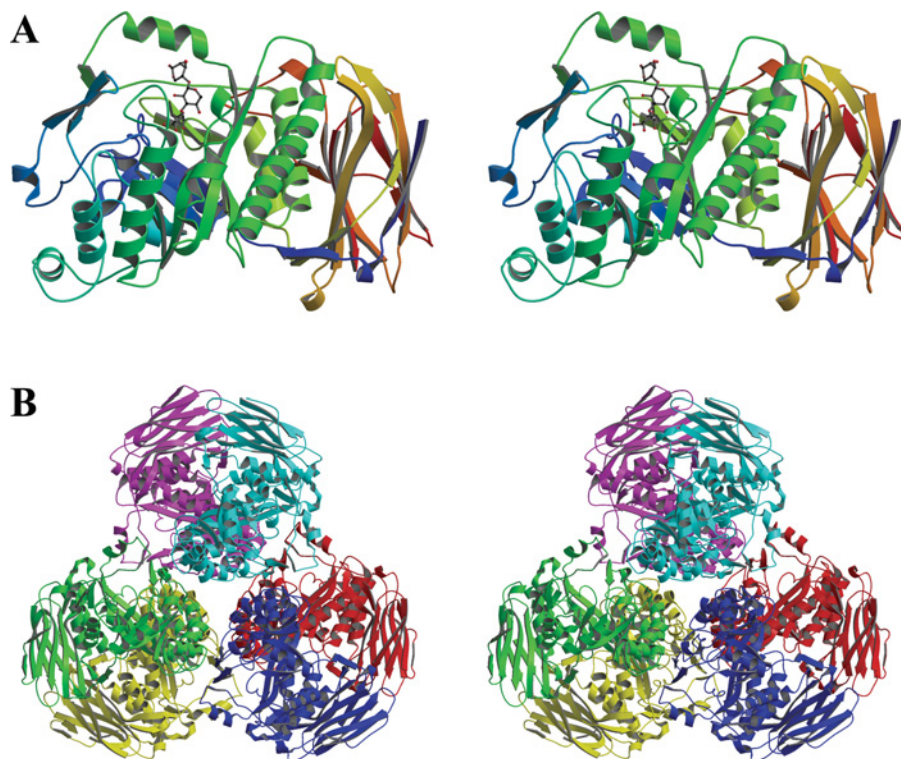


Figure 2 The three-dimensional structure of *CtAraf51*

(A) Divergent (wall-eyed) stereo representation of a monomer of *CtAraf51* with the ligand observed in the AX2 complex shown in ball-and-stick representation. The cartoon is 'colour-ramped' from the N-terminus (blue) to C-terminus (red). (B) A hexamer of *CtAraf51* with each monomer in a different colour.

were dissolved adjacent to the crystals and soaked between 5 and 30 min. A similar addition of ligand was made to the cryo-solution prior to bathing. All data were collected at the ESRF (European Synchrotron Radiation Facility; Grenoble, France) on station ID14-3. Single crystals were maintained at 100 K and data were collected with an oscillation width of 0.2° over a 120° range, at a wavelength of 0.9310 \AA ($1 \text{ \AA} = 0.1 \text{ nm}$), using a MAR 165 CCD (charge-coupled-device) detector.

Structure solution and refinement

All X-ray data were processed and integrated using MOSFLM [18] and scaled using SCALA with other computing using the CCP4 suite [19], unless stated otherwise. *CtAraf51A* crystals were found to belong to space group $P4_12_12$ or $P4_32_12$ with the approximate cell dimensions of $a = b = 173.5 \text{ \AA}$, $c = 272.2 \text{ \AA}$. Native data to approx. 3.1 \AA , but of poor quality, were used to produce an initial partial model that was used in subsequent refinement of the complex crystal structures. The native *CtAraf51A* structure was solved by molecular replacement using the program PHASER with a monomer from the structure of the *G. stearothermophilus* Araf51 (PDB code 1QW9) as a search model. The space group was confirmed to be $P4_32_12$ and six molecules were found in the asymmetric unit. Five per cent of the observations were set aside for cross-validation and used to monitor the progress of the refinement and for appropriate weighting of geometrical and temperature factor restraints. The program COOT [20] was used to make manual corrections to the model between cycles of refinement using the program REFMAC [21]. The complex structures were validated using PROCHECK [22]. Figures 2 and 3 were drawn with PyMOL (DeLano Scientific,

San Carlos, CA, U.S.A.; <http://pymol.sourceforge.net/>) or MOLSCRIPT/BOBSCRIPT [23,24].

RESULTS AND DISCUSSION

Activity of *CtAraf51*

The gene encoding *CtAraf51* was expressed in *E. coli* cells at a high level and the full-length protein could be purified, a procedure eased by a simple heat denaturation step by virtue of the heat stability of the clostridial enzyme. Kinetics was performed on a variety of natural and non-natural substrates (Table 1). *CtAraf51* is, unsurprisingly, a thermophilic enzyme with $T_{\text{opt}} \sim 82^\circ\text{C}$ with PNP-Araf as substrate. Not surprisingly, the enzyme is most active on the artificial substrate PNP-Araf ($k_{\text{cat}}/K_m \sim 400 \text{ s}^{-1} \cdot \text{mM}^{-1}$ measured at 37°C), which contains a relatively good leaving group ($\text{p}K_a$ of 4-nitrophenyl is 7.2). *CtAraf51* was also highly active on small soluble oligosaccharides, notably α -1,5-linked arabinobiose/arabinotriose and the AX2 which reflect the natural limit dextrin products of the action of arabinanases on arabinan and xylanases on wheat arabinoxylan. Indeed, *CtAraf51* is also highly efficient in the removal of the α -1,3-linked arabinoside decorations of polymeric wheat arabinoxylan itself, although it is unlikely that this intracellular enzyme would encounter intact xylan *in vivo*. *CtAraf51* displays no significant activity against linear non-substituted arabinan, which likely reflects the low concentration of available arabinofuranoside moieties (the non-reducing terminal sugar is $< 0.2\%$ of the substrate) or xyloglucan. Surprisingly, *CtAraf51* exhibits very low activity against sugar beet arabinan, and is thus unable to access the O-3 and O-2 linked arabinofuranose side chains in this polysaccharide. The activity of this clostridial enzyme is thus notably different from some

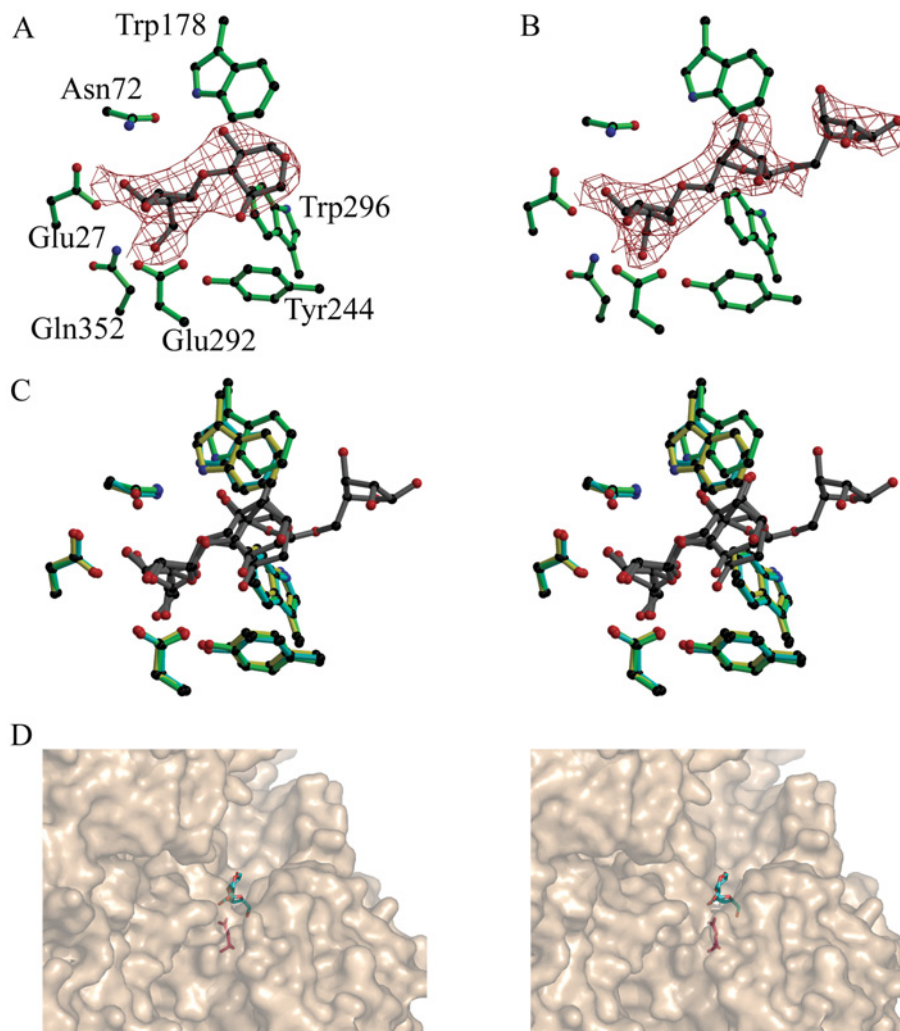


Figure 3 The active centre of *CtAraf51*

Observed electron density for the AX2 complex (**A**) and A3 complex (**B**) of *CtAraf51* with interacting residues shown at the electron density contoured at approx. 1σ . (**C**) The Figure shows the overlap of the two complexes of *CtAraf51* (AX2 grey; A3 green) with the AX complex (yellow) of the *G. stearothermophilus* enzyme [14] in divergent (wall-eyed) stereo. The protein surface (**D**) reveals that the specificity site for the arabinoside moiety is at the bottom of a funnel-shaped depression which permits accommodation of the terminal arabinoside derived from different oligosaccharides with almost equal catalytic efficiency (the nucleophile is shown in red liquorice and the Ara-Xyl from the AX2 complex in cyan).

Table 1 Catalytic activity of *CtAraf51* on a panel of different substrates

ND, not detectable.

Substrate	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m ($s^{-1} \cdot mM^{-1}$)
PNP-Araf	0.25 ± 0.03	103 ± 12.9	412
Wheat arabinoxylan	0.0055 ± 0.001	0.07 ± 0.009	12.7
AX2	0.95 ± 0.2	8.6 ± 0.9	9.1
A2	0.55 ± 0.001	4.6 ± 0.3	8.4
Sugar beet arabinan*	ND	ND	0.5
Linear arabinan	ND	ND	ND
Xyloglucan	ND	ND	ND

* Although enzyme activity was present, it was not possible to dissect out the individual kinetic constants.

previously well-characterized enzymes such as the Araf51 from *Cellvibrio japonicus* which is much more efficient in removing the side chain decorations from xylan and arabinan than in hydrolysing α -1,5-linked arabinofuranose substrates derived from

arabinan [11]. Furthermore, the likely locations of these enzymes are different. *CtAraf51* displays no signal peptide and is thus likely to be intracellular, whereas the *Ce. japonicus* enzyme is believed to be membrane-bound [4]. The specificity and location of *CtAraf51* are discussed, in light of the three-dimensional structure, below.

Structure of *CtAraf51*

CtAraf51 was crystallized and could be solved by molecular replacement using the related *G. stearothermophilus* enzyme [14], with which *CtAraf51* shares approx. 65% sequence identity, as the search model. The 503 residues of *CtAraf51* can be traced from residue Lys² through Gly⁵⁰² with no breaks, with the exception of the A3 complex where residues 483–485 of chain E are disordered.

Model building and refinement were non-trivial at the resolution of 2.7–2.9 Å and were aided greatly by the averaging of the six copies of *CtAraf51* in the asymmetric unit and the application of TLS (translation, libration and screw-rotation) refinement as implemented in REFMAC [25]. Crystallographic *R* values

Table 2 X-ray data and structure quality statistics for *CtAraf51*

R.M.S.D., root mean square deviation.

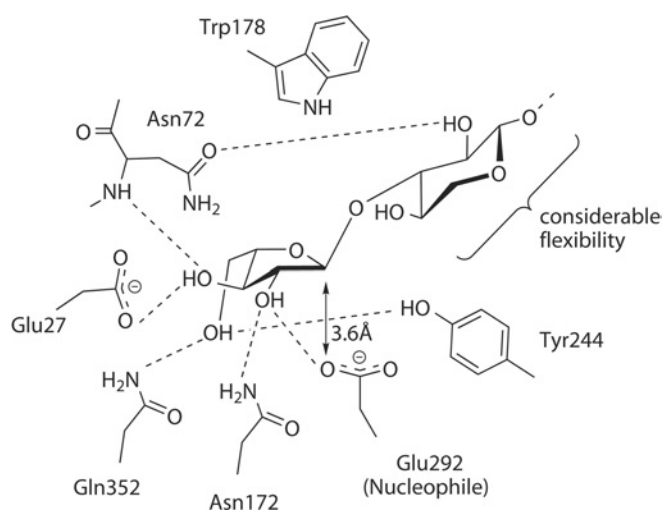
Structural data	GH51 AX2 complex	GH51 A3 complex
Resolution (outer shell) (Å)	63.37–2.9 (3.06–2.90)	59.76–2.70 (2.85–2.70)
<i>R</i> merge	0.118 (0.567)	0.098 (0.586)
Mean <i>I</i> / σ <i>I</i>	9.8 (2.3)	11.2 (2.8)
Completeness (%)	99.1 (99.6)	99.9 (99.9)
Multiplicity (outer shell)	3.8 (3.8)	3.7 (3.8)
No. of unique reflections	91349	113914
No. of protein atoms	24068	23743
No. of ligand atoms	144	186
No. of solvent waters	233	117
<i>R</i> _{cryst}	0.20	0.21
<i>R</i> _{free}	0.25	0.28
R.M.S.D. 1–2 bonds (Å)	0.007	0.017
R.M.S.D. bond angles (°)	1.34	1.99
R.M.S.D. chiral volume (Å ³)	0.010	0.15
Mean main chain B (Å ²)	53	46
Mean side chain B (Å ²)	53	46
Mean substrate B (Å ²)	68	55
Mean solvent B (Å ²)	43	39
Ramachandran statistics		
% Allowed region	97.9	97.8
PDB code	2C8N	2C7F

and deviations from stereochemical target values are as one would expect at this resolution (Table 2). As with the *G. stearothermophilus* Araf51, the clostridial monomer displays an unusual architecture. The catalytic domain is an elaborated (β/α)₈ barrel, approximately comprising residues 24–389, which is followed by a 12-stranded β -sandwich domain one of whose strands is formed by the N-terminal 14 residues of the enzyme (Figure 2). As with the *G. stearothermophilus* Araf51, the enzyme is arranged as a functional hexamer both in crystal and in solution (light scattering; results not shown) with 32 point group symmetry (a trimer of dimers).

Complex structures and the catalytic mechanism of *CtAraf51*

Family GH51 enzymes perform catalysis with net retention of anomeric configuration through the formation, and subsequent breakdown of a covalent β -L-linked arabinofuranosyl-enzyme intermediate [26]. Two essential catalytic residues are required for such a reaction, an acid/base and an enzymatic nucleophile. The structure of the (β/α)₈ barrel domain places GH51 into the superfamily defined as clan GH-A [15] in which both catalytic functions are glutamates that are found on strands β -4 (acid/base) and β -7 (nucleophile) respectively. In the case of *CtAraf51*, these residues are Glu¹⁷³ and Glu²⁹² respectively. In order to allow complex formation, an inactive acid/base variant, Glu¹⁷³Ala, was constructed and shown to be inactive on natural substrates possessing poor leaving groups, such as A3 and AX2.

Three-dimensional complexes were obtained with both A3 and AX2 (Table 2). Both complexes are of the 'Michaelis' type, with unhydrolysed ligand spanning the active centre (Figure 3) and thus with an arabinofuranoside moiety in subsite -1 (nomenclature according to [27]). The active centre is a funnel-shaped canyon, with specificity for arabinofuranosides in the -1 subsite (discussed below), but with sufficient breadth 'above' to allow accommodation of different shapes and sizes of aglycone. An appropriate aromatic residue in the +1 subsite, Trp¹⁷⁸, forms hydrophobic 'stacking' with the aglycone moiety, xylose or arabinose in the AX2 and A3 complexes respectively, and both the side chain of this residue and the main chain of the 176–179

**Figure 4** Interactions in the active centre of *CtAraf51*

Hydrogen bonds < 3.2 Å are shown as dotted lines.

loop show conformational flexibility when accommodating AX2 versus A3. This open funnel is furnished with a number of other aromatic/hydrophobic residues with only a single direct hydrogen bond to the O-2 hydroxyl of xylose in the AX2 complex and none at all to the aglycone regions of A3 in the +1 subsite or beyond. All other ligand hydroxy groups hydrogen-bond only with solvent water molecules, which presumably accounts for the relative promiscuity of *CtAraf51* compared with some of the more specific GH51 members. The topography of the active centre, in which the ligand points down into the funnel, also clearly permits the binding of decorated xylans (and arabinans) in which the arabinofuranoside moiety is perpendicular to the main hemi-cellulose chain. Indeed, in the AX2 complex, density beyond the first xylose group is both disordered and extending in both 'directions'. This suggests that there is little specificity for the exact orientation of the +1 xyloside, and thus the enzyme could accommodate arabinofuranosides in extended xylo-oligosaccharides (Figure 4) as one might expect for an enzyme with high catalytic efficiency on arabinoxylan (Table 1).

In the 'catalytic' -1 subsite, the arabinoside in the AX2-derived complex lies in an *E*₃ (envelope) conformation and that from A3 in ⁴*T*₀ (twist-boat) conformation with the consequence that the leaving group C-1–O glycosidic bond is axial with the nucleophile and with Glu²⁹² poised for in-line nucleophilic attack, as observed on other related Michaelis complexes [14,28–30]. The OE2 to C-1 distance is 3.6 Å. The O-2 of the arabinofuranoside moiety interacts both with the nucleophile OE1 atom (2.6 Å) and with Asn¹⁷² of the 'NE' (where 'E' is the catalytic acid/base) signature motif of many clan GH-A enzymes although in the case of *CtAraf51*, the third member of this motif is not a proline, as more normally observed, but methionine. The O-3 interacts both with the side chain of Glu²⁷ and the main chain amide of Asn⁷² with the adjacent side chain of Phe⁷³ providing the steric blockage that forms the base of the funnel and which contributes to the exo-catalytic activity. The O-5 of the exocyclic hydroxymethyl group is sandwiched between Gln³⁵² (NE2 at 3.0 Å) and Tyr²⁴⁴ (OH at 3.0 Å). The interactions of the -1 subsite observed here, are similar to those reported for the *G. stearothermophilus* enzyme [14]. The only difference is that the sugar ring conformation observed in the *G. stearothermophilus* enzyme was a ⁴*T*₀ twist conformation, whereas we have modelled the conformation on the *Cl. thermocellum* enzyme as an *E*₃ envelope. This difference

may simply reflect the lower resolution of the data available for *CtAraf51*.

Summary

The substrate specificity of GH51 enzymes reflects the different mechanisms by which micro-organisms utilize plant structural polysaccharides. For example, the GH51 arabinofuranosidases of *Ce. japonicus* and *Streptomyces chartreusis* are approx. 300-fold more active against α -1,2- and α -1,3-linked arabinofuranose residues than α -1,5 arabino-oligosaccharides. This is consistent with the observation that the Gram-negative bacterium utilizes arabinotriose but not arabinofuranose as a growth substrate, and thus the hydrolysis of the trisaccharide by the major extracellular arabinofuranosidase, *CjAraf51A*, would have a detrimental effect on the survival of the prokaryote. In contrast, *CtAraf51* displays considerable plasticity with respect to the α -glycosidic bond and the nature of the aglycone component of the substrate. This is entirely consistent with the role of the intracellular clostridial enzyme which is likely to be the release of arabinose moieties from xylan-derived decorated oligosaccharides and linear α -1,5-linked arabino-oligosaccharides. While the substrate specificity of the *G. stearothermophilus* enzyme against substrates other than aryl-glycosides is unknown, its cytoplasmic location and high sequence identity with *CtAraf51* suggest that the *Geobacillus* hydrolase displays a similar role and specificity to the clostridial glycoside hydrolase. This proposal is entirely consistent with the observation that Trp¹⁷⁸, which is the key specificity determinant for the aglycone sugar in *CtAraf51*, is conserved in the *Geobacillus* arabinofuranosidase as is the flexible loop extending from 176 to 179.

The complete degradation of plant cell wall polysaccharides requires the orchestrated action of a complex enzymatic consortium. In the case of xylan degradation, 'side chain' decorations such as glucuronosyl, acetyl, arabino and feruloyl moieties must be removed either prior or subsequent to the action of endo-xylanases, and prior to the action of β -xylosidases. In many cases, the accessory enzymes will be secreted, or remain intimately associated with the membrane. In the case of the cellulosome-possessing *Cl. thermocellum*, the vast majority of enzymes involved in plant cell wall degradation will be attached, through their dockerin domains (described at the structural level in [31]), to the complex multienzyme machine termed the cellulosome. There is increasing awareness that some enzymes, somewhat counter-intuitively, reside inside the cell. Examples include the *G. stearothermophilus* α -glucuronidase [32], although this enzyme species only targets xylo-oligosaccharides in which the non-reducing xylose is decorated with an α -1,2-linked 4-*O*-methylglucuronic acid moiety irrespective of its cellular location [33], and some acetyl-xylo-oligosaccharide deacetylases (discussed in [34]). The absence of any signal peptide on the *CtAraf51* also demands that this enzyme is intracellular, suggesting an emerging picture that xylan breakdown products are not simply imported into organisms as mono/disaccharides, but as partially substituted sugars retaining their arabinoside, acetyl and glucuronyl substituents. Such a role may account for the relaxed specificity of the *CtAraf51* reported in the present study.

We thank the Biotechnology and Biological Sciences Research Council and the Royal Society for funding and the staff of the ESRF for provision of data collection facilities. We also thank Dr Mark Fogg (York) and the EU SPINE project for their assistance with ligation-independent cloning.

REFERENCES

- Saha, B. C. (2000) Alpha-L-arabinofuranosidases: biochemistry, molecular biology and application in biotechnology. *Biotechnol. Adv.* **18**, 403–423
- Gilead, S. and Shoham, Y. (1995) Purification and characterization of alpha-L-arabinofuranosidase from *Bacillus stearothermophilus* T-6. *Appl. Environ. Microbiol.* **61**, 170–174
- Sheehan, J. and Himmel, M. (1999) Enzymes, energy, and the environment: a strategic perspective on the US Department of Energy's Research and Development Activities for Bioethanol. *Biotechnol. Prog.* **15**, 817–827
- Beylot, M. H., Emami, K., McKie, V. A., Gilbert, H. J. and Pell, G. (2001) *Pseudomonas cellulosa* expresses a single membrane-bound glycoside hydrolase family 51 arabinofuranosidase. *Biochem. J.* **358**, 599–605
- Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316
- Henrissat, B. and Davies, G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644
- Davies, G. J., Gloster, T. M. and Henrissat, B. (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Curr. Opin. Struct. Biol.* **15**, 637–645
- Flipphi, M. J., Visser, J., van der Veen, P. and de Graaff, L. H. (1994) Arabinase gene expression in *Aspergillus niger*: indications for coordinated regulation. *Microbiology* **140**, 2673–2682
- Nurizzo, D., Turkenburg, J. P., Charnock, S. J., Roberts, S. M., Dodson, E. J., McKie, V. A., Taylor, E. J., Gilbert, H. J. and Davies, G. J. (2002) A novel 5-blade β -propeller with enzymatic activity revealed by the 3-D structure of the *Cellobivibrio cellulosa* processive α -L-arabinanase, Arb43A. *Nat. Struct. Biol.* **9**, 665–668
- Tateishi, A., Mori, H., Watari, J., Nagashima, K., Yamaki, S. and Inoue, H. (2005) Isolation, characterization, and cloning of α -L-arabinofuranosidase expressed during fruit ripening of Japanese pear. *Plant Physiol.* **138**, 1653–1664
- Beylot, M. H., McKie, V. A., Voragen, A. G. J., Doeswijk-Voragen, C. H. L. and Gilbert, H. J. (2001) The *Pseudomonas cellulosa* glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity. *Biochem. J.* **358**, 607–614
- McKie, V. A., Black, G. W., Millward-Sadler, S. J., Hazlewood, G. P., Laurie, J. I. and Gilbert, H. J. (1997) Arabinanase A from *Pseudomonas fluorescens* subsp. *cellulosa* exhibits both an endo- and an exo-mode of action. *Biochem. J.* **323**, 547–555
- Matsuo, N., Kaneko, S., Kuno, A., Kobayashi, H. and Kusakabe, I. (2000) Purification, characterization and gene cloning of two α -L-arabinofuranosidases from *Streptomyces chartreusis* GS901. *Biochem. J.* **346**, 9–15
- Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. and Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 alpha-L-arabinofuranosidase. *EMBO J.* **22**, 4922–4932
- Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P. and Davies, G. (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 7090–7094
- Bayer, E. A., Belaich, J. P., Shoham, Y. and Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* **58**, 521–554
- Proctor, M., Taylor, E. J., Nurizzo, D., Turkenburg, J. P., Lloyd, R. M., Vardakou, M., Davies, G. J. and Gilbert, H. J. (2005) Tailored catalysts for plant cell-wall degradation: redesigning the exo/endo preference of the *Cellobivibrio japonicus* arabinanase 43A. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 2697–2702
- Leslie, A. G. W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESRF-EACMB Newsletter on Protein Crystallography, vol. 26, Daresbury Laboratory, Warrington, U.K.
- Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132
- Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum likelihood method. *Acta Crystallogr. D* **53**, 240–255
- Laskowski, R. A., McArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 282–291
- Kraulis, P. J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950
- Esnouf, R. M. (1997) An extensively modified version of MolScript that includes greatly enhanced colouring capabilities. *J. Mol. Graph. Model.* **15**, 132–134
- Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D* **57**, 122–133
- Pitson, S. M., Voragen, A. G. J. and Beldman, G. (1996) Stereochemical course of hydrolysis catalyzed by arabinofuranosyl hydrolases. *FEBS Lett.* **398**, 7–11

- 27 Sulzenbacher, G., Shareck, F., Morosoli, R., Dupont, C. and Davies, G. J. (1997) The *Streptomyces lividans* family 12 endoglucanase: construction of the catalytic core, expression, and X-ray structure at 1.75 Å resolution. *Biochemistry* **36**, 16032–16039
- 28 Davies, G. J., Mackenzie, L., Varrot, A., Dauter, M., Brzozowski, A. M., Schulein, M. and Withers, S. G. (1998) Snapshots along an enzymatic reaction coordinate: analysis of a retaining beta-glycoside hydrolase. *Biochemistry* **37**, 11707–11713
- 29 Ducros, V., Zechel, D. L., Murshudov, G. N., Gilbert, H. J., Szabo, L., Stoll, D., Withers, S. G. and Davies, G. J. (2002) Substrate distortion by a β -mannanase: snapshots of the Michaelis and covalent intermediate complexes suggest a B_{2,5} conformation for the transition-state. *Angew. Chem. Int. Ed. Engl.* **41**, 2824–2827
- 30 Davies, G. J., Ducros, V. M.-A., Varrot, A. and Zechel, D. L. (2003) Mapping the conformational itinerary of β -glycosidases by X-ray crystallography. *Biochem. Soc. Trans.* **31**, 523–527
- 31 Carvalho, A. L., Dias, F. M. V., Prates, J. A. M., Nagy, T., Gilbert, H. J., Davies, G. J., Ferreira, L. M. A., Romão, M. J. and Fontes, C. M. G. A. (2003) Cellulosome assembly revealed by the crystal structure of the cohesin–dockerin complex: a template for nanosomal catalytic machines. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13809–13814
- 32 Shulami, S., Gat, O., Sonenshein, A. L. and Shoham, Y. (1999) The glucuronic acid utilization gene cluster from *Bacillus stearothermophilus* T-6. *J. Bacteriol.* **181**, 3695–3704
- 33 Nurizzo, D., Nagy, T., Gilbert, H. J. and Davies, G. J. (2002) The structural basis for catalysis and specificity of the *Pseudomonas cellulosa* α -glucuronidase, GlcA67A. *Structure* **10**, 547–556
- 34 Vincent, F., Charnock, S. J., Verschuere, K. H. G., Turkenburg, J. P., Scott, D. J., Offen, W. A., Roberts, S., Pell, G., Gilbert, H. J., Davies, G. J. et al. (2003) Multifunctional xylooligosaccharide/cephalosporin C deacetylase revealed by the hexameric structure of the *Bacillus subtilis* enzyme at 1.9 Å resolution. *J. Mol. Biol.* **330**, 593–606

Received 3 November 2005/1 December 2005; accepted 5 December 2005

Published as BJ Immediate Publication 5 December 2005, doi:10.1042/BJ20051780